# Semi-Supervised Distance Metric Learning for Person Re-Identification

Feng Chen and Jinhong Chai
Logistic Service Center
Yunnan Power Grid Co., LTD
Kunming ,China
AC2873@163.com
369998250@qq.com

Dinghu Ren
Kunming Dongdian Technology Co., LTD
Kunming ,China
dinghu_ren@ddtech.com.cn

Xiaofang Liu and Yun Yang
School of Software
Yunnan University
Kunming ,China
18716023506@163.com
yangyun@ynu.edu.cn

*Abstract*— **As a fundamental task in automated video surveillance, person re-identification, which has received increasing attention in recent years, aims to match people across non-overlapping camera views in a multi-camera surveillance system. It has been reported that KISS metric learning has been followed by most of the previous supervised work because of its state of the art performance for person re-identification on VIPeR dataset. However, given only a small number of labeled image pairs available for training, the matching model certainly suffers from unstable learning process and poor matching result. To address this serious practical issue, we proposed a novel semi-supervised KISS metric learning (SS-KISS) approach which makes use of unlabeled data to improve the re-identification performance by 1) combining both global and local information to select the most confident image pairs from the unlabeled data; 2) using an ensemble approach, which explores advantages of supervised and unsupervised learning by reconciling two matching models on which labeled and unlabeled data to an optimal one via smart weighting schema. Extensive experiments have been conducted on three datasets: VIPeR, ETHZ, and i-LiDS, experimental results demonstrate that our approach achieves a sound performance in the case of small amount of labeled data.**

***Keywords—ensemble learning; metric learning; person re-identification; semi-supervised learning***

## I. Introduction

Person Re-identification handles pedestrian recognizing and associating at different locations and time after who had been previously observed visually within multiple camera networks with non-overlapping field-of-view. Extensively, re-identification technology has a wide range of practical applications in material procurement and warehouse management in the aspects of security assurance and video surveillance by saving a lot of resources on exhaustively retrieving a target of interest from a large amount of video sequences. As a result, this field has drawn a rapid attention from both academic researchers and industrial developers. Although many approaches have been proposed to improve the performance of person re-identification system from different perspectives, researchers still face many challenges and unsolved problems in the real-world situation: (a) crowed public scene with a number of possible matched pedestrians, (b) significant visual appearance changes due to varying lighting conditions, viewing angles, body poses, background clutter and occlusions across camera views. To address these challenges, researchers have to focus on (1) feature-based method for extracting discriminative features for robust representation, and (2) metric learning based method for developing more accurate matching models.

Compared with feature-based method, the metric learning based method, which intends to match certain probe images against a gallery of person in another camera view, has less requirement on feature representation and can obtain greater accuracy of recognition, thus has been particularly prized than feature-based method. Matching models are generally categorized into unsupervised [2-6] and supervised methods [7-11]. Unsupervised methods mainly concern the design and extraction of robust visual features and do not involve additional human labeling efforts. Most of the top ranked state of-the-art feature sets employ regional or patch-based features, while others improve the performance by combining different types of features. For example, symmetry-driven accumulation of local features (SDALF)[2], custom pictorial structures (CPS)[3], biologically inspired features and covariance descriptors (BiCov)[4], local descriptors encoded by Fisher vectors combined with other features (eLDFV)[5],and salient dense correspondence combined with other features (eSDC) [6]. Otherwise a method is regarded as a supervised approach if prior to application, it exploits labeled samples for tuning model parameters such as distance metrics, feature weight or decision boundaries. Common works include KISS metric learning (KISS) [12], Large Margin Nearest Neighbors (LMNN)[9], Information Theoretic Metric Learning (ITML)[10], Logistic Discriminant Metric Learning (LDML)[13] and PCCA[14], Rank Support Vector Machines (RankSVM)[8] etc. By taking the advantages of both supervised and unsupervised methods, more recently, some semi-supervised learning works for re-identification [15-19] and other fields[20, 21] have been proposed and could be worthy of reference for person re-identification.

Although the supervised approaches generally have better performance with the assistance of manually labeled training

Natural Science Foundation of China (NSFC) under grant number: 61402397 and 61663046.

Applied Basic Research Foundation of Yunnan Province under the grant number: 2016FB104.

Applied Basic Research Foundation of Yunnan Province under the grant number: 2016FB104

Program for Yunnan High Level Overseas Talent Recruitment.

Scientific Research Innovation Foundation of Yunnan University Graduate Student.

978-1-5090-5507-4/17/$31.00 ©2017 IEEE
ICIS 2017, May 24-26, 2017, Wuhan, China

IEEE computer society

samples, in real-world situation, where the number of cameras is large, it is impractical to label training samples for every pair of camera views, which significantly affects the performance of a supervised learning strategy to train a robust model. Thus, some supervised models especially based on Gaussian distribution with limited training data always suffer from estimation error or over-fitting. Yet, unsupervised approaches eliminate the dependence on pre-annotated training data, and thus do not require any manual labeling, but they always get a poor results than supervised approaches. Many efforts have been made to address the challenges of re-identification in both directions. In fact, few of works have been carried out to synthesize the best of both words effectively. In this paper, we therefore propose an original semi-supervised ensemble method by selecting potential image pairs from the unlabeled data and proposing weighting scheme which evaluates quality of models learned from labeled and unlabeled data via KISS distance metric, and combining multiple matching results come out of two models into an optimal solution.

The main contributions of this paper are summarized as follows: First, we propose a new method to select the most confident potential image pairs from the unlabeled data by combining global and local information for semi-supervised learning in person re-identification. Thus, using only a small number of labeled data and a large number of selected higher prior unlabeled data for training to overcome the fundamental weakness under few available labeled data. Second, we develop a semi-supervised ensemble algorithm with a novel weighting scheme, which optimally reconciles matching results based on supervised and unsupervised models into a single consolidated solution, where the weight definition does not involve human judgment, but an automated learning process. Finally, we demonstrate the effectiveness and the efficiency of our approach on the most popular and recognized benchmark datasets, and experimental results show that our approach performs well.

The rest of this paper is organized as follows. Section 2 describes the details of our proposed model and its relevant theoretical analysis. Section 3 reports the simulations on VIPeR, ETHZ, and i-LIDS benchmarks and discusses the issues related to our approach. Finally, the conclusion is drawn in Section 4.

## II. Our Approach

In this section, we present our semi-supervised learning model as a whole and review the underpinning techniques used in our simulations followed by a specific description for our semi-supervised metric learning approach.

### A. Model Description

The proposed approach consists of two major modules; i.e. parameters learning module based on KISS and semi-supervised ensemble matching module which reconciles matching results come out of supervised and unsupervised models into a single one via a weighting scheme. As illustrated in Fig. 1, the procedure of implementing our approach can be described as follows:

1) In the parameters learning module, KISS metric learning, which is presented in part B, is used to learn the supervised matching model from labeled training data firstly. Then, potential image pairs are selected from unlabeled data by the method presented in part C to learning the unsupervised matching model. Finally, a novel weighting scheme is proposed to assess the importance of two models according to its discriminative power, where weights are intrinsically derived from the labeled training set, and larger weight should be assigned to better quality matching model on the corresponding dataset.

2) In the semi-supervised ensemble of matching module, several matching results are determined by matching specified probe images or tracks against a gallery of persons in another camera view using supervised and unsupervised matching models from different datasets, then these matching results are finally combined into an optimal solution via a novel weighting scheme, which is described in part C in detail.

### B. Review of KISS Metric Learning

As a distance metric learning approach, the KISS (Keep it simple and straight) metric learning has obtained the state of the art performance for person re-identification on the VIPeR dataset [22]. From a statistical inference point of view the optimal statistical decision whether a pair $(\mathrm{x_i}, \mathrm{x_j})$ is dissimilar or not can be obtained by a likelihood ratio test. Thus, we test the hypothesis $H_0$ that a pair is dissimilar versus the alternative $H_1$:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log\left(\frac{\mathrm{p}(\mathbf{x}_i, \mathbf{x}_j|\mathrm{H_0})}{\mathrm{p}(\mathbf{x}_i, \mathbf{x}_j|\mathrm{H_1})}\right) \quad (1)$$

A positive value of $(\mathbf{x}_\mathrm{i}, \mathbf{x}_j)$ means that $H_0$ is validated. Incontrast, a low value means that $H_0$ is rejected and the pairis considered as similar. To be independent of the actual locality in the feature space, we cast the problem in the space of pairwise differences($\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$) with zero mean and can re-write Eq. (1) to:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log\left(\frac{\mathrm{p}(\mathbf{x}_i, \mathbf{x}_j|\mathrm{H_0})}{\mathrm{p}(\mathbf{x}_i, \mathbf{x}_j|\mathrm{H_1})}\right) = \log\left(\frac{f(\mathbf{x}_{ij}|\theta_0)}{f(\mathbf{x}_{ij}|\theta_1)}\right) \quad (2)$$

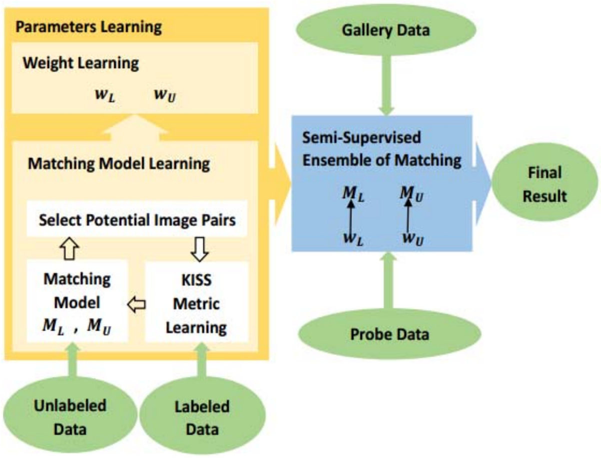

Fig. 1. Semi-supervised distance learning model

Where $f\left(\mathbf{x}_{ij}|\theta_1\right)$ is a probability density function with parameters $\theta_1$ for hypothesis $H_1$ that a pair $\left(\mathbf{x}_\mathrm{i}, \mathbf{x}_j\right)$ is similar ($y_{ij} = 1$) and vice-versa $H_0$ for a pair being dissimilar. It is common to assume that $\mathbf{x}_{ij}$ satisfies the Gaussian distribution. Thus, we can relax the problem and rewrite Eq. (2) to:

$$\delta\left(\mathbf{x}_i, \mathbf{x}_j\right) = \log\left(\frac{\frac{1}{\sqrt{2\pi\left|\Sigma_{y_{ij}=0}\right|}} exp\left(-\frac{1}{2}\mathbf{x}_{ij}^T\Sigma_{y_{ij}=0}^{-1}\mathbf{x}_{ij}\right)}{\frac{1}{\sqrt{2\pi\left|\Sigma_{y_{ij}=1}\right|}} exp\left(-\frac{1}{2}\mathbf{x}_{ij}^T\Sigma_{y_{ij}=1}^{-1}\mathbf{x}_{ij}\right)}\right) \tag{3}$$

the covariance matrices are estimated as follows:

$$\Sigma_{y_{ij}=1} = \frac{1}{N_1}\Sigma_{y_{ij}=1}\left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \tag{4}$$

$$\Sigma_{y_{ij}=0} = \frac{1}{N_0}\Sigma_{y_{ij}=0}\left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \tag{5}$$

Where $N_0$ denotes the number of dissimilar vector pairs and $N_1$ denotes the number of similar vector pairs. By taking the log, we can reformulate the likelihood test as:

$$\delta\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{x}_{ij}{}^T\left(\Sigma_{y_{ij}=1}^{-1} - \Sigma_{y_{ij}=0}^{-1}\right)\mathbf{x}_{ij} \tag{6}$$

### C. SS-KISS Metric Learning

Although KISS metric learning had improved person re-identification, there is plenty of scope to improve its stability and efficiency. Specifically, it is known that the Gaussian distribution model suffers from estimation error when given limited samples and finally results in poor performance. Therefore, we proposed the semi-supervised ensemble learning approach, which attempts to train a better classification model by incorporating a small amount of labeled data with a large amount of unlabeled data.

The key component is how to fully make the best of the available information in data. Selecting potential positive and negative data pairs from unlabeled training data by combining global and local information such as mentioned in [23] is the first step which can obtain some potential and meaningful information from huge amount of training data. Practically, we divided our training data into two parts: labeled training data set $\mathrm{X}_L$ and unlabeled training data set $\mathrm{X}_U$.Since the matching model can be holly ascertained by $\mathrm{M} = \Sigma_{y_{ij}=1}^{-1} - \Sigma_{y_{ij}=0}^{-1}$ according to Equation(6). Conveniently, let $\mathrm{M}_L = \Sigma_{y_{ij}=1}^{-1} - \Sigma_{y_{ij}=0}^{-1}$ denote supervised matching model if $\mathbf{x}_i$ and $\mathbf{x}_j$ come from labeled training data set $\mathrm{X}_L$ and $\mathrm{M}_U = \Sigma_{\bar{y}_{ij}=1}^{-1} - \Sigma_{\bar{y}_{ij}=0}^{-1}$ denote unsupervised matching model if $\mathbf{x}_i$ and $\mathbf{x}_j$ come from unlabeled training data set $\mathrm{X}_U$.where, $\bar{y}_{ij} = 1$ means we regard unlabeled data $\mathbf{x}_i$ and $\mathbf{x}_j$ as an similar pair. Conversely, $\bar{y}_{ij} = 0$ means an dissimilar pair. To improve the performance of Basic models $\mathrm{M}_L$ and $\mathrm{M}_U$, we adopt regularized smoothing technology mentioned in[1] to preprocess the model. One way to determine the positive pairs is to label the potential positive image pairs with very high scores, i.e., classify $\mathrm{k}_1$ most closest unknown examples into a class using point $\mathbf{x}_i$ as the center based on their KISS distance, and thus potential positive images set $\mathbf{x}_\mathrm{p} = \{\mathbf{x}_i, \mathbf{x}_{j1}, \mathbf{x}_{j2} \ldots, \mathbf{x}_{j\mathrm{k}_1}\}$, we combine any two of them together as an similar pair $\bar{y}_{ij} = 1$. As for negative pairs, we select $\mathrm{k}_2$ examples randomly from unlabeled data set when $\delta(\mathbf{x}_i, \mathbf{x}_{hk}) > \mathbf{x}_m$, and potential negative images set $\mathbf{x}_n = \{\mathbf{x}_i, \mathbf{x}_{h1}, \mathbf{x}_{h2} \ldots, \mathbf{x}_{h\mathrm{k}_2}\}$, we combine any two of $\mathbf{x}_j$ and $\mathbf{x}_h$ together as an dissimilar pair $\bar{y}_{ij} = 0$. Where, $\{\mathbf{x}_i, \mathbf{x}_{j1}, \mathbf{x}_{j2} \ldots, \mathbf{x}_{j\mathrm{k}_1}\} \cup \{\mathbf{x}_i, \mathbf{x}_{h1}, \mathbf{x}_{h2} \ldots, \mathbf{x}_{h\mathrm{k}_2}\} \subset \mathrm{X}_U$.

$$\mathbf{x}_m = \frac{1}{n}\sum_{k=1}^{n}\delta(\mathbf{x}_i, \mathbf{x}_k) \tag{7}$$

Finally, the final matching result can be obtain by:

$$\begin{aligned}\delta^*\left(\mathbf{x}_i, \mathbf{x}_j\right) &= w_L\mathbf{x}_{ij}{}^T\mathrm{M}_L\mathbf{x}_{ij} + w_U\mathbf{x}_{ij}{}^T\mathrm{M}_U\mathbf{x}_{ij} \\ &\overset{\text{def}}{=} w_L\delta_L\left(\mathbf{x}_i, \mathbf{x}_j\right) + w_U\delta_U\left(\mathbf{x}_i, \mathbf{x}_j\right)\end{aligned} \tag{8}$$

Where,$w_L$ and $w_U$ mean weight of supervised and unsupervised models respectively. In our approach, larger weight should be assigned to the better matching result obtained on the more discriminative model. Thus, the discriminative power of model could be determined by a ratio between inter-person distance and intra-person distance on labeled training set as follows:

$$w_L = \frac{\varDelta_L}{\varDelta_L + \varDelta_U} \tag{9}$$

$$\Delta_\tau = \frac{\Sigma_{y_{ij}=0}\left(\delta_\tau(\mathbf{x}_{ij}) + \left|min\left(\delta_\tau(\mathbf{x}_{ij})\right)\right|_{y_{ij}=1}\right)}{\Sigma_{y_{ij}=1}\left(\delta_\tau(\mathbf{x}_{ij}) + \left|min\left(\delta_\tau(\mathbf{x}_{ij})\right)\right|_{y_{ij}=1}\right)} \tag{10}$$

Where, a high value $\Delta_\tau$ means that the model makes a large distance and a small distance, therefore have a more discriminative power. And of course $w_U = 1 - w_L$ .

### D. Algorithm Design

```
Algorithm Training procedure of SS-KISS
Input: labeled training data set $\mathrm{X}_L$ and unlabeled training data set $\mathrm{X}_U$.
1) Learn supervised KISS model $M_L$ from $\mathrm{X}_L$ by equation(6);
2) Select confidence potential similar images pairs $\bar{y}_{ij} = 1$ and dissimilar images pairs $\bar{y}_{ij} = 0$ from $\mathrm{X}_U$;
3) Learn unsupervised model $M_U$ from data set selected by step 2;
4) Learn weights on $\mathrm{X}_L$ by (9) and (10);
5) Learn final matching model by (8);
Output: Optimal matching model $\delta^*$
```

## III. Simulation

### A. Datasets and settings

Three publically available person re-identification datasets, VIPeR[20] ETHZ [7], and i-LIDS Multiple-Camera Tracking Scenario (MCTS)[24, 25] were used for evaluating the performance of our approach. The widely used VIPeR dataset includes 1,264 outdoor images of 632 subjects taken from two different views with normalized size at 128 × 64 pixels. Viewpoint, illumination and pose changes were the most prominent cause of matching challenge. The ETHZ dataset was originally designed for pedestrian detection and tracking captured from a moving camera in a busy street scene and then modified for person re-identification dataset. There are 8580 images of 146 people with normalized size at 128 × 64 pixels in total. The number of detections per person ranges from 5 to 356. The challenges of this dataset are the illumination changes and occlusions on people's appearance whilst the view angle change is small. In the i-LIDS MCTS dataset, which was captured indoor at a busy airport arrival hall, there are 119 people with a total 476 person images captured by multiple non-overlapping cameras with an average of 4 images for each person. The images were normalized to a size of 128 × 64 pixels. Many of these images undergo large illumination change, considerable view angle change, and are subject to large occlusions. It is noted that these three datasets have different characteristics (e.g. outdoor/indoor, large/small variations in view angle, presence/absence of occlusion) and therefore are ideal for evaluating person re-identification algorithms given different challenges. Among them, the ETHZ dataset is considered to be the easiest one due to the fact that it was not actually captured by multiple non-overlapping view cameras and thus lack of view angle change. Note that across the three datasets, the average number of training images of each person ranges from 2 (VIPeR) to 6 (ETHZ) highlighting the under-sampled class distribution typical for the person re-identification problem.

In our experiments, we randomly selected all images of $p$ people to set up the training set, and the rest people were used for testing. During training, all images of p/2 people were selected randomly to set up the labeled training set $\mathrm{X}_L$, and all images of rest p/2 people to set up the unlabeled training set $\mathrm{X}_U$ defined in Sec. II. For the labeled training set, a pair of images of each person formed a positive pair, and one image of him/her and one of another person in the training set formed a negative pair. While for the unlabeled training set, positive pairs and negative pairs formed by the scheme in part C of section II. Note that, the number of labeled pairs and unlabeled pairs are equal. Each test set was composed of a gallery set and a probe set. The gallery set consisted of one image for each person, and the remaining images were used as the probe set.

### B. Feature Representation

We apply our SS-KISS model as well as other models to an appearance representation of people captured by a set of different basic features. Specifically, the image representation is the combination of RGB, YCbCr, and HSV color features and two texture features extracted by Schmid and Gabor filters[26] on six horizontal strip, and totally 13 Schimid filters and 8 Gabor filters were obtained. In total 29 feature channels were constructed for each stripe and each feature channel was represented by a 16 dimensional histogram vector. Similar to those employed by[7, 8], each person image was thus represented by a feature vector in a 2784 dimensional feature space. Since the features computed for this representation include low-level features widely used by existing person re-identification techniques, this representation is considered as generic and representative..

### C. Baselines and Performance Measures

We first compared our approach with five representative metric learning approaches to validate the effectiveness of our approach, including Relative Distance Comparison (RDC)[11], Adaboost[27], Bhattacharyya distance(Bhat)[28], Partial Least Squares (PLS)[29] , and Xing's[30]. Each of these methods has its own merits, and some of them shown the state of the art performance in many applications. Furthermore, we compared our semi-supervised ensemble KISS approach with KISS approach disposed by regularized smoothing technology under different training data situations to reveal the superiority of our approach.

For evaluation, we use the Cumulative Matching Characteristic (CMC) curve [11, 31, 32], which is the standard performance measurements for this task. A rank $r$ matching rate indicates the percentage of the probe images with correct matches found in the top $r$ ranks against the gallery images. Note that in practice, as the gallery size increases, it becomes more difficult to find the correct match and CMC curves become lower.

### D. Experiment Results

In our experiments, we conduct semi-supervised ensemble technique to learn a robust distance metric and measure the distance between two images and judge whether or not the two images come from a same person.

TABLE I provides the results compared to state-of-the-art methods on different datasets: VIPeR on the left, ETHZ on the middle and i-LIDS on the right. It is worth mentioning that our semi-supervised ensemble model using only half of labeled data(all images of p/2 subject) and other of people unlabeled. Yet, the results show that semi-supervised ensemble metric

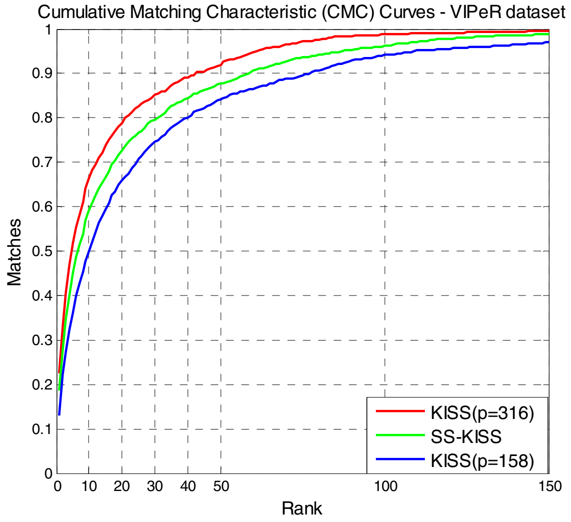

Fig. 2. Performance on the VIPeR dataset in term of CMC curves. In each subfigure, the x-coordinate is the rank score and y-coordinate is the matching rate. Only the top 150 ranking positions are depicted.

TABLE I. PERSON RE-IDENTIFICATION TOP MATCHING RATES ON VIPER, ETHZ, AND I-LIDS DATASET: COMPARING WITH THE POPULAR ALGORITHMS

| Method | VIPeR(P=316) | | | | ETHZ(p=106) | | | | i-LIDS(p=76) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *r=1* | *r=10* | *r=25* | *r=50* | *r=1* | *r=5* | *r=10* | *r=20* | *r=1* | *r=5* | *r=10* | *r=20* |
| Our Approch | **0.186** | **0.589** | **0.767** | 0.878 | 0.711 | 0.889 | 0.953 | **0.992** | 0.312 | 0.570 | 0.686 | 0.851 |
| RDC | 0.157 | 0.539 | 0.752 | 0.879 | **0.727** | **0.901** | **0.956** | 0.988 | **0.441** | **0.727** | **0.847** | **0.963** |
| Adaboost | 0.082 | 0.366 | 0.582 | **0.909** | 0.692 | 0.878 | 0.935 | 0.980 | 0.356 | 0.664 | 0.799 | 0.932 |
| Bhat | 0.047 | 0.166 | 0.266 | 0.402 | 0.610 | 0.809 | 0.878 | 0.941 | 0.318 | 0.614 | 0.742 | 0.895 |
| PLS | 0.027 | 0.109 | 0.204 | 0.329 | 0.546 | 0.751 | 0.833 | 0.924 | 0.258 | 0.574 | 0.736 | 0.903 |
| Xing's | 0.047 | 0.166 | 0.266 | 0.415 | 0.608 | 0.803 | 0.874 | 0.936 | 0.303 | 0.626 | 0.773 | 0.906 |

Some results are directly taken from[1] .

learning can obtain almost same performance as advanced supervised learning approach which has been limited in real application on VIPeR and ETHZ dataset. Admittedly, our approach does poorly on i-LIDS dataset because a small amount of unlabeled data is available. In summary, when the number of training samples is insufficient, SS-KISS achieves satisfying matching accuracy with assistance of enough unlabeled data.

In Fig.2, Fig.3 and Fig.4, we compare the proposed SS-KISS with regularized smoothing KISS on VIPeR, ETHZ and i-LIDS datasets in different conditions respectively. We randomly selected p and p/2 labeled people to form the training set for KISS respectively. By contrast, we randomly selected p/2 labeled and p/2 unlabeled people to set up the training set for SS-KISS. The main observations from the matching performance comparisons are given below.

1) Figures show that the simply added labeled samples can improve the matching accuracy. However, seen another way, supervised learning approach has too weak discriminative power when labeled training data is limited, as the blue curve shown.
2) The proposed SS-KISS improves KISS to a comparable performance by taking full advantage of the underlying information in unlabeled data under the situation when the number of training samples is insufficient. And thus avoids the drawbacks of covariance matrix estimation error in KISS when given a small size of training data.

It is not surprising to observe that the supervised learning-based KISS approaches outperform our semi-supervised approach. In general, supervised learning approaches are dominant, not surprisingly, for person re-identification and outperform unsupervised learning approaches. Nevertheless, the semi-supervised approaches benefit from unlabeled data and finally reconciled with labeled data by a weight schema. The results suggest that supervised and unsupervised method are not exclusive, but can complement each other to improve re-identification accuracy.

## IV. CONCLUSION

In this work, a semi-supervised ensemble learning framework has been proposed to deal jointly with the supervised-based and unsupervised-based re-identification problem using only a small quantity of labeled image pairs and some unlabeled image pairs. The novel semi-supervised KISS metric learning approach adds the most confident potential image pairs from the unlabeled data by combining both global and local information for training to improve the re-identification performance. Furthermore, we achieve a convincing performance by using an ensemble approach, which explores advantages of su-

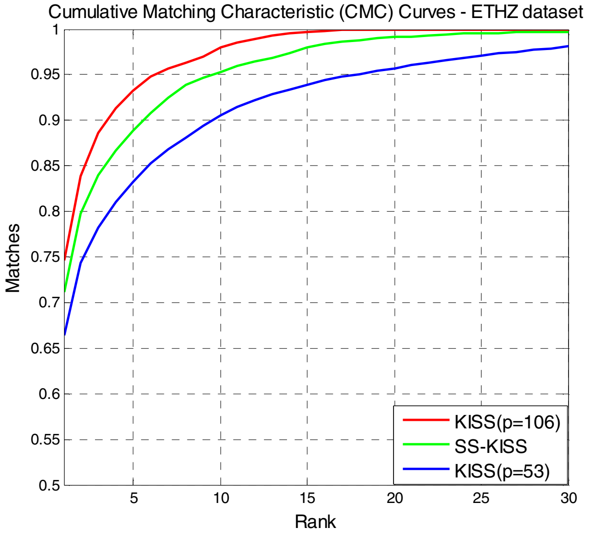

Fig. 3. Performance on the ETHZ dataset in term of CMC curves. In each subfigure, the x-coordinate is the rank score and y-coordinate is the matching rate. Only the top 30 ranking positions are depicted.

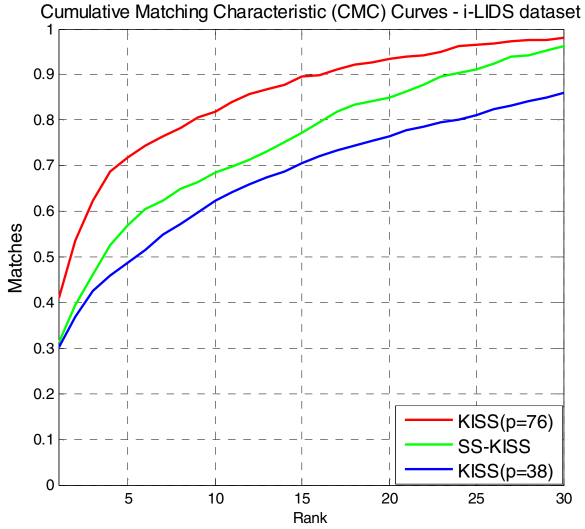

Fig. 4. Performance on the i-LIDS dataset in term of CMC curves. In each subfigure, the x-coordinate is the rank score and y-coordinate is the matching rate. Only the top 30 ranking positions are depicted.

pervised and unsupervised learning by reconciling two matching models on which labeled and unlabeled data to an optimal one via smart weighting schema.

Experiments on different benchmarks demonstrate significant improvement made by our algorithm using only a small number of labeled image pairs for training in comparison to the baseline and state-of-the-art supervised algorithms. In our ongoing work, we are seeking an automatic alternative optimization strategy for unlabeled data selecting process in a more effective way and exploring potential real applications.

## ACKNOWLEDGEMENT

Authors are grateful to acknowledge the financial support from Natural Science Foundation of China (NSFC) under grant number: 61402397 and 61663046. Applied Basic Research Foundation of Yunnan Province under the grant number: 2016FB104, Yunnan Key Laboratory of software engineering General Program under grant： 2015SE201, Scientific Research Foundation for the Returned Overseas Scholars, Program for Yunnan High Level Overseas Talent Recruitment, and Scientific Research Innovation Foundation of Yunnan University Graduate Student.

## REFERENCES

[1] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing kiss metric learning," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 23, pp. 1675-1685, 2013.

[2] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," Computer Vision and Image Understanding, vol. 117, pp. 130-144, 2013.

[3] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom Pictorial Structures for Re-identification," in BMVC, 2011, p. 6.

[4] B. Ma, Y. Su, and F. Jurie, "Bicov: a novel image representation for person re-identification and face verification," in British Machive Vision Conference, 2012, p. 11 pages.

[5] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey, "A database for person re-identification in multi-camera surveillance networks," in Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on, 2012, pp. 1-8.

[6] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3586-3593.

[7] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in Computer Vision–ECCV 2008, ed: Springer, 2008, pp. 262-275.

[8] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person Re-Identification by Support Vector Ranking," in BMVC, 2010, p. 6.

[9] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," The Journal of Machine Learning Research, vol. 10, pp. 207-244, 2009.

[10] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in Proceedings of the 24th international conference on Machine learning, 2007, pp. 209-216.

[11] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in Computer vision and pattern recognition (CVPR), 2011 IEEE conference on, 2011, pp. 649-656.

[12] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 2288-2295.

[13] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in Computer Vision, 2009 IEEE 12th international conference on, 2009, pp. 498-505.

[14] A. Mignon and F. Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 2666-2672.

[15] M. Bauml, M. Tapaswi, and R. Stiefelhagen, "Semi-supervised learning with constraints for person identification in multimedia data," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3602-3609.

[16] D. Figueira, L. Bazzani, H. Q. Minh, M. Cristani, A. Bernardino, and V. Murino, "Semi-supervised multi-feature learning for person re-identification," in Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on, 2013, pp. 111-116.

[17] U. Iqbal, I. D. Curcio, and M. Gabbouj, "Who is the hero? semi-supervised person re-identification in videos," in Computer Vision Theory and Applications (VISAPP), 2014 International Conference on, 2014, pp. 162-173.

[18] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu, "Semi-supervised coupled dictionary learning for person re-identification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3550-3557.

[19] A. J. Ma and P. Li, "Semi-Supervised Ranking for Re-identification with Few Labeled Image Pairs," in Asian Conference on Computer Vision, 2014, pp. 598-613.

[20] Y. Yang and X. Liu, "A robust semi-supervised learning approach via mixture of label information," Pattern Recognition Letters, vol. 68, pp. 15-21, 2015.

[21] Y. Yang, Z. Li, W. Wang, and D. Tao, "An adaptive semi-supervised clustering approach via multiple density-based information," Neurocomputing, 2017.

[22] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), 2007.

[23] Y. Yang and J. Jiang, "Hybrid sampling-based clustering ensemble with global and local constitutions," IEEE transactions on neural networks and learning systems, vol. 27, pp. 952-965, 2016.

[24] W.-S. Zheng, S. Gong, and T. Xiang, "Associating Groups of People," in BMVC, 2009, p. 6.

[25] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35, pp. 653-668, 2013.

[26] Y. Zhang and S. Li, "Gabor-LBP based region covariance descriptor for person re-identification," in Image and Graphics (ICIG), 2011 Sixth International Conference on, 2011, pp. 368-371.

[27] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using haar-based and dcd-based signature," in Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, 2010, pp. 1-8.

[28] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE transactions on communication technology, vol. 15, pp. 52-60, 1967.

[29] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on, 2009, pp. 322-329.

[30] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," in NIPS, 2002, p. 12.

[31] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in Computer Vision–ECCV 2012, ed: Springer, 2012, pp. 780-793.

[32] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1-8.